



Co-funded by the Prevention of and Fight
against Crime Programme of the European Union

RESPONSE PROJECT DATABASE - NPS AND RELATED COMPOUNDS

DESCRIPTION and GUIDELINES FOR USE

DRAFT DOCUMENT

Dr. Sonja Klemenc

RS Ministry of the Interior, General Police Directorate
National Forensic Laboratory



Ljubljana, December 2016

Table of contents

1. Foreword	3
2. Disclaimer	4
3. Database description and guidelines for use	5
3.1 Examples of use	9
4. Substance descriptors: structure, common name, systematic name, standard InChI key and consistency of systematic chemical identifiers within and between databases.....	16
4.1 RESPONSE project conversions of structure to systematic name and opposite, standard InChI Key and cross validations strategy.....	18
5. References.....	21



Co-funded by the Prevention of and Fight
against Crime Programme of the European Union

This work has been produced with the financial support of the Prevention of and fight against crime Programme of the European Union (grant agreement number JUST/2013/ISEC/DRUGS/AG/6413). The contents of this publication are the sole responsibility of the author and can in no way be taken to reflect the views of the European Commission.

1. Foreword

The main goal of the RESPONSE¹ NPS part of the project was to provide and share numerous MS and FTIR spectral data on newly appearing or recently reported NPS.

Several sources of materials have been characterized and results implemented into database. Undoubtedly certified reference materials (CRM) are the most reliable source. The drawback is that CRMs are very expensive and beside this they are not available in real time, i.e. when laboratory detects a new compound. The producers are usually several steps - months behind the "gray" internet market. Thus other sources of NPS have been explored: NPS materials purchased over internet (test purchases), seized samples (police, customs), samples collected anonymously (in Slovenia) or samples collected from other laboratories.

The selection of analytical methods for chemical characterizations of samples categories listed above is dependent on the sample type and prior information available about the sample/ compound. In general the following instrumental methods have been applied: GC-MS (EI ionization 70 eV), HPLC-TOF, FTIR-ATR, GC-(MS)-IR (condensed phase), NMR (1D and 2D experiments) and later on ion chromatography (anions) and in some specific cases melting point measurements were applied as well. Interpretations of experimentally obtained data and correct identification of particular compound are checked by at minimum two experts independently. Some interpretations are supported by cheminformatic tools. In addition, from the beginning of 2016, the structure conformation/ elucidation of the test purchased samples by NMR partner is "blind", i.e. it is done from the base. Data provided to NMR analysts are: exact monoisotopic mass, empirical formula, information of salt or base form of compound and solubility data.

Already in the first year the Slovenian partners (National Forensic Laboratory (NFL) and Faculty of chemistry and Chemical Technology (FKKT), University Ljubljana) identified many novel compounds and generated numerous MS, IR and NMR spectra. Therefore, to share gathered analytical information promptly, which means almost in real time, the public open »DRUGS MONOGRAPHS – NPS and related compounds"² was developed and launched at the project web site at the end of 2015. Database is searchable along several criteria and at the same time it supports identification of »unknown« substance if its mass spectrum (electron ionization at 70eV) is known.

For great majority of compounds “Analytical reports” have been issued and implemented into database. The first part of the Analytical report describes the sample (reference material, seized, collected or test purchased sample). For reference materials specific information which guarantees traceability back to the vendor and a lot number is given. In the next session of the report analytical parameters followed by supporting experimental data obtained during chemical characterization of particular sample are given. The types and quantity of supporting experimental data depend on sample type and previous information already available from other resources (databases and/or scientific publications). In 2016 database was enriched with several reports from the contributors outside the RESPONSE project. We kindly acknowledge this and at the same time encourage the new contributors to participate!

*Dr. Sonja Klemenc
RESPONSE
project manager*

2. Disclaimer

Although “RESPONSE” analytical results are reviewed independently by at minimum two experts of National Forensic Laboratory (NFL) chemistry department and NMR interpretations by two experts from Faculty of Chemistry and Chemical Technology, University of Ljubljana, it is still sole responsibility of database user to critically re-evaluate our findings. In the case of observed inconsistencies, i.e. possible identification errors please report them to sonja.klemenc@policija.si.

RESPONSE database shall be regarded as supporting tool only and not as an officially peer-reviewed database.

3. Database description and guidelines for use

Database window is shown on Figure 1. Database is structured as a table. Each row is a record which describes one compound. Heading row currently has 18 fields – compound/sample descriptors - see details in the Table 1). Each field has a *sort button* (example is labeled by red circle no 1 on Figure 1) which enables sorting of entries (alphabetically or ascending/descending order).

On the top left side of the window there is a field “show” (red circle no. 2 on Figure 1). It serves for the adjustments of database window, i.e. how many records one wish to see per page. Options are: 5, 10, 25, 50 and All.

At the bottom (left corner) there is information of what you see on the screen (example on Figure 1: *Showing 1 to 5 of 422 entries*), a date of the last database update is given below and counter of database “visits” is at the bottom.

Buttons on the bottom right side of the window enables movements between the pages. Scroll bar enables left – right movements along the database window. Vertical scroll bar will appear on as needed basis.

Substance class	substance (NPS) common name	structure (created by OPSIN free tool)	NPS1 systematic name	other names	Formula per base form	Mw (g/mol) per base form NPS1	MS (BP1)	MS (BP2)	MS (BP3)	GC-MS-RT NFL/min	MS spectrum (picture)	StdIn ChIKey	Type of detection	comments	date of entry	report updated	Chem-ID or Report-ID
Aminoindanes	N-Methyl-2AI		N-methyl-2,3-dihydro-1H-inden-2-amine	NM-2AI, N-methyl-2-aminoindane	C10H13N	147.22	147	42	106	2.73	show	S0U7QUC1TORHJL-UHFFFAOYSA-N	test purchase		2015-10-25	2015-12-03	1349-15
Aminoindanes	MDAl		2H,5H,6H,7H-indeno[5,6-d][1,3]dioxol-6-amine	5,6-methylenedioxy-2-aminoindane	C10H11NO2	177.20	160	149	177	4.31	show	F0DRMHCLVZAXJM-UHFFFAOYSA-N	test purchase		2015-11-25	2015-12-03	1347-15
Aminoindanes	2AI		2,3-dihydro-1H-inden-2-ylamine	2-Aminoindane	C9H11N	133.19	133	116	91	2.55	show	LWHHFZAXSANGGM-UHFFFAOYSA-N	RM-reference material	I-SEE project RM and report	2016-01-20		1469-16
Aminoindanes	5IAI		5-iodo-2,3-dihydro-1H-inden-2-amine	5-iodo-2-aminoindane	C9H10IN	259.09	259	115	117	4.86	show	RHPYCDOPGHVQO-UHFFFAOYSA-N	RM-reference material	I-SEE project RM and report	2016-01-20		1475-16
Aminoindanes	2AI		2,3-dihydro-1H-inden-2-amine	2-Aminoindane	C9H11N	133.19	133	116	91	2.50	show	LWHHFZAXSANGGM-UHFFFAOYSA-N	RM-reference material		2016-07-18		1598-16

Showing 1 to 5 of 422 entries
Table updated 28.10.2016
8727 hits

Previous 1 2 3 4 5 ... 85 Next

scroll bar

Figure 1: Database window

On the top right side of the window (red circle no. 3 on Figure 1) there is a field “search”. You can search for any reasonable parameter. The search field functions as a filter.

Table 1: Database fields description and guidelines for use

	Field	description	Comments/actions-use	notes
1	Substance class	Substances are classified into several classes along their physicochemical characteristics / <i>aminoindanes, arylalkylamines, arylcyclohexylamines, benzodiazepines, cannabinoids, cathinones, indolalkylamines (tryptamines), opioids, others, plants & extracts.</i> This classification was adopted from EMCDDA. ³	Click the sort button – records will be sorted from A-Z or in opposite direction; By typing the name of particular class into the search field you can select all records from particular class and then use the sort buttons in other fields for deeper search.	
2	Substance Common name	Be aware that this is not a unique name (for the same substance we can find more than one common name). Basically we (NFL) follow EMCDDA naming or apply other names available over the internet or we create the name in the NFL if the compound is reported for the first time.	Click the sort button – records will be sorted from A-Z or in opposite direction; Click on the link (compound name) to see the “ Analytical report ” (pdf format)	
3	Structure	Picture of structure (in .png format)	Click or double click to enlarge – structure opens in new window.	Structure – name –structure is crosschecked by free tools available over the internet (procedures are explained later on).
4	NPS1 systematic name	Systematic name – IUPAC based name which describes compound structure	Click the sort button to sort records ascending/ descending order. Apply the search field to search for the whole name or only a part of the name (for example indole or indazole or carboxylate..)	In the RESPONSE project data base name is always given per base form of compound. Structure – name –structure is crosschecked by free tools available over the internet (procedures are explained later on).
5	Other names	Other names for compound which are available over the internet at the time of the compound entry (IUPAC based or common names).	Click the sort button – to sort records ascending/ descending order. Apply the search field to search for the whole or only a part of the name (for example AKB.or PICA or PINACA, etc).	
6	Formula per base form	Empirical formula of the compound per base form.	Click the sort button to sort the records (ascending/descending) along the number of C atoms.	
7	M _w per base form (g/mol)	Molecular weight per base form.	Click the sort button to sort the records ascending/descending).	

	Field	description	Comments/actions-use	notes
8	GC-MS-RT NFL (min)	Retention time (RT) as measured in the National forensic laboratory (NFL).	Click the sort button to sort the records.	In NFL we use the "retention time locked" GC-MS method. Therefore, absolute RT shows long term reproducibility at our analytical conditions. For other user this field is not very useful.
9	MS (BP1)	The most intensive peak (m/z) in the mass spectrum.	Click the sort button to sort the records (ascending/ descending order).	For NFL data the peak areas calculated automatically are taken into account. Be aware that isotopic peaks are taken into account. As well.
10	MS (BP2)	Second most intensive peak (m/z) in the mass spectrum.	Click the sort button to sort the records (ascending/ descending order)	
11	MS (BP3)	Third most intensive peak (m/z) in the mass spectrum.	Click the sort button to sort the records (ascending/ descending order)	
12	MS spectrum (picture)	Mass spectrum (EI, +70eV) picture (in .png format).	Click to see the picture which opens in new window	Majority of MS spectra were collected in NFL. Exception is the spectra provided from contributors of analytical reports (see type of detection field).
13	StdInChIKey	Standard InChI Key = standardized International Chemical Identifier is a textual identifier for small organic molecules. The condensed, 27 characters standard InChIKey is a hashed version of the full standard InChI string, designed to allow for easy web searches of chemical compounds. ⁴	Click the link and search the internet for information on particular compound.	Code is crosschecked by free tools available over the internet (see heading 4.1).
14	Type of detection	<p>the source of compound (sample) or analytical information sources:</p> <p><u>Analytical report</u> – analytical data were provided from contributors outside NFL.</p> <p><u>Collected</u> – samples collected anonymously in Slovenia or provided from other laboratories</p> <p><u>In house RM</u> – in house reference materials obtained in the frame of proficiency tests or assessment of the lab</p> <p><u>RM-reference materials</u> – purchased from official vendors</p> <p><u>Seized</u> – samples seized by the Slovenian police.</p> <p><u>Test purchase</u> – purchased samples which are advertised and sold over the internet as legal highs or research chemicals (not reference materials!)</p>	<p>Click the sort button to sort the records alphabetically;</p> <p>By typing the name of particular "type of detection" into the search field you will see all records of the same type of detection. Afterwards you can use the sort buttons in other fields for deeper search.</p>	Analytical strategy is dependent on the type of detection and is described in a separate document. (see below)
15	Comments	comments about the analytical findings and/or contributor of the sample or analytical report	Click the sort button to sort ascending/descending.	* in the future (when database structure will be improved) this field will be spitted into two fields: analytical comments and separate field for contributors
16	Date of entry	date when the compound was entered into database	Click the sort button to sort ascending/descending	

	Field	description	Comments/actions-use	notes
17	Report updated	Date of the Analytical report update	Click the sort button to sort ascending/descending	Some "cosmetic" improvements of the reports will not be labeled as updates.
18	Chem-ID or Report-ID	Chem-ID is the unique label of particular sample in the NFL chemicals database. REPORT-ID is the unique label of the analytical report provided by contributors outside NFL.	Click the sort button to sort ascending/descending	This code guarantees the traceability to the sample and/or original source of chromatographic, spectroscopic and other data provider for the particular sample.

DRAFT

3.1 Examples of use

Example 1: The SEARCH field

One is interested only on reference materials records, sorted by molecular weight.

Procedure:

- in the "search" field type "reference"
- apply "sort" button in M_w (g/mol) field

Database respond as obtained at 31/10/2016: 222 entries (filtered from 422 total entries), sorted in ascending values of M_w (Figure 2).

To remove the current filter click on blue cross in the search field or apply the F5 key on the keyboard.

Search: ✕

Substance class	substance (NPS) common name	structure (created by OPSIN free tool)	NPS1 systematic name	other names	Formula per base form	M_w (g/mol) per base form NPS1	MS (BP1)	MS (BP2)	MS (BP3)	GC-MS-RT NFL/min	MS spectrum (picture)	StdIn ChIKey	Type of detection	comments	date of entry
Others	Indane		2,3-dihydro-1H-indene	1,2-hydrindene, hydrindonaphthene, hydrindene	C9H10	118.18	117	118	115	2.09	show	PQHFJRRNORRO:UHFFFAQYSA-N	RM-reference material		2016-08-03
Aminoindanes	2-AI		2,3-dihydro-1H-inden-2-ylamine	2-Aminoindane	C9H11N	133.19	133	116	91	2.55	show	LWHEZAKSANGGM:UHFFFAQYSA-N	RM-reference material	I-SEE project RM and report	2016-01-20
Aminoindanes	2-AI		2,3-dihydro-1H-inden-2-amine	2-Aminoindane	C9H11N	133.19	133	116	91	2.50	show	LWHEZAKSANGGM:UHFFFAQYSA-N	RM-reference material		2016-07-18
Others	1,3-Indanedione		2,3-dihydro-1H-indene-1,3-dione	1,3-Dioxindan	C9H6O2	146.15	146	104	76	2.90	show	LHKAJLSQFADET:UHFFFAQYSA-N	RM-reference material		2016-08-03
Aminoindanes	N-methyl-2AI		N-methyl-2,3-dihydro-1H-inden-2-amine	2,3-dihydro-N-methyl-1H-inden-2-amine NMe-2AI N-methyl-2-Aminoindane	C10H13N	147.22	147	42	106	2.75	show	SOJZOUJTORHJT:UHFFFAQYSA-N	RM-reference material		2016-07-18

Showing 1 to 5 of 222 entries (filtered from 422 total entries)

Table updated 28.10.2016
18737 hits

Previous: 1 2 3 4 5 ... 45 Next

Figure 2: Filtered records on reference materials and sorted by M_w .

Example 2: Identify your "unknown" by searching RESPONSE project database MS spectral data (BP1, BP2, BP3) and picture of MS spectrum. This example comes from the real request.

Let's assume that an expert has the mass spectrum as shown on Figure 3 in hand. Mw is not known, the compound class is not known.

First step:

Read peak intensities from your spectrum: $BP1=144$, $BP2=144$, $BP3=233$

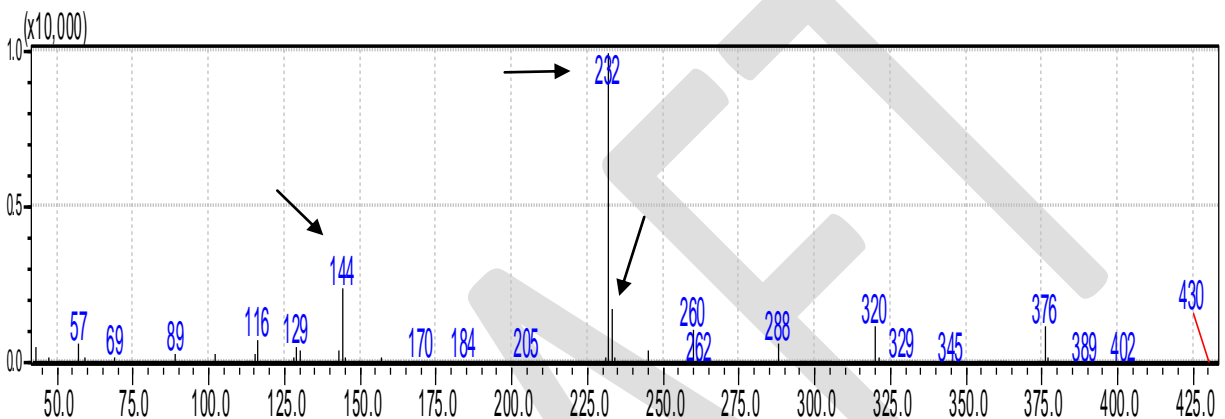


Figure 3: Mass spectrum of the "unknown" compound (real request)

Second step:






Search the RESPONSE project database:

- Clear all previous selections from the "search field"
- In the "search field" type: 232 144 233 (put spaces between numbers)



Figure 4: Search field criteria

- Result (as obtained at 20/11/2016) is shown in Figure 5: *Six compounds* had been filtered out of 422 entries.

Substance class	substance (NPS) common name	structure (created by OPSIN free tool)	NPS1 systematic name	other names	Formula per base form	Mw (g/mol) per base form NPS1	MS (BP1)	MS (BP2)	MS (BP3)	GC-MS-RT NFL/min	MS spectrum (picture)	StdIn ChIKey
Cannabinoids	5F-MDMB-PICA		methyl 2-[[1-(5-fluoropentyl)-1H-indol-3-yl]formamido]-3,3-dimethylbutanoate	5F-MDMB-2201, MDMB-2201	C21H29FN2O3	376.46	232	144	233		show	CH5UEEBESACQDV-UHFFFAQYSA-N
Cannabinoids	5F-PB-22		quinolin-8-yl 1-(5-fluoropentyl)-1H-indole-3-carboxylate	5-fluoro QUPIC, 5-fluoro-PB-22	C23H21FN2O2	376.43	232	144	233	17.11	show	MBOCMBFDVVSGLJ-UHFFFAQYSA-N
Cannabinoids	5-fluoro-JWH-018-adamantyl analog		3-(adamantane-1-carbonyl)-1-(5-fluoropentyl)-1H-indole	[1-(5-fluoropentyl)-1H-indol-3-yl]tricyclo[3.3.1.13.7]dec-1-yl-methanone; 5F-AB-001; AM2201 adamantyl analog	C24H30FNO	367.51	232	144	233	15.37	show	V/BLLMFXPGHZFFP-UHFFFAQYSA-N
Cannabinoids	LTI-701		1-(5-fluoropentyl)-N-phenyl-1H-indole-3-carboxamide		C20H21FN2O	324.40	232	144	233	14.05	show	FZYJALPV/LNXEIP-UHFFFAQYSA-N
Cannabinoids	LTI-701		1-(5-fluoropentyl)-N-phenyl-1H-indole-3-carboxamide		C20H21FN2O	324.40	232	144	233	14.01	show	FZYJALPV/LNXEIP-UHFFFAQYSA-N

Showing 1 to 6 of 6 entries (filtered from 422 total entries)

[Previous](#)[Next](#)

Table updated 28.10.2016

19748 hits

Figure 5: Filtered entries for the criteria given in search field

- d. To inspect the mass spectra of filtered compounds and compare them with your data click the “show” button in the column “MS spectrum picture” (picture opens in new window).
- e. RESULT: For this particular example the mass spectrum of 5F-MDMB-PICA (Figure 6) was in good agreement with the spectrum in question.

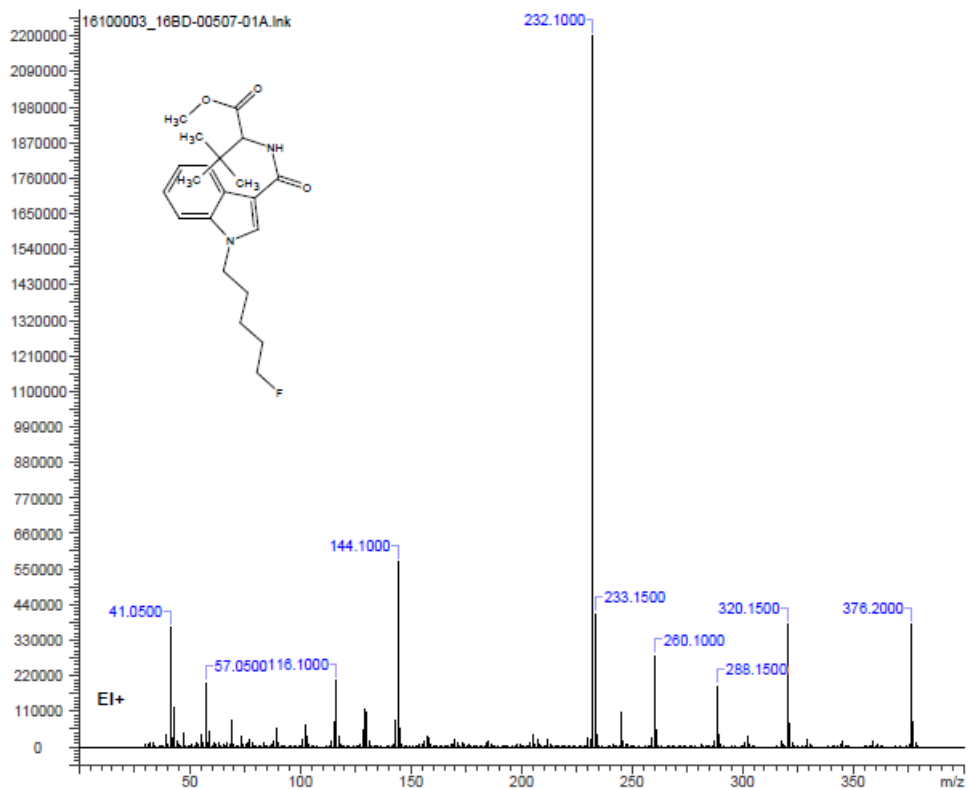







Figure 6: 5F-MDMB-PICA spectrum in the RESPONSE project database

Third step:

See »Analytical report« (click the compound common name link) to see if some other analytical data are presented (as for example FTIR and other analytical data).

To search the WEB for information on this substance – click the StdInChI Key link.

Substance class	substance (NPS) common name	structure (created by OPSIN free tool)	NPS1 systematic name	other names	Formula per base form	Mw (g/mol) per base form NPS1	MS (BP1)	MS (BP2)	MS (BP3)	GC-MS-RT NFL/min	MS spectrum (picture)	StdIn ChiKey
Cannabinoids	5F-MDMB-PICA		methyl 2-[[1-(5-fluoropentyl)-1H-indol-3-yl]formamido]-3,3-dimethylbutanoate	5F-MDMB-2201, MDMB-2201	C21H29FN2O3	376.46	232	144	233		show	CH5UEEBESACQDV-UHFFFAQYSA-N
Cannabinoids	5F-PB-22		quinolin-8-yl 1-(5-fluoropentyl)-1H-indole-3-carboxylate	5-fluoro QUPIIC, 5-fluoro-PB-22	C23H21FN2O2	376.43	232	144	233	17.11	show	MBOCMBFDYVSGLJ-UHFFFAQYSA-N
Cannabinoids	5-fluoro-JWH-018-adamantyl analog		3-(adamantane-1-carbonyl)-1-(5-fluoropentyl)-1H-indole	[1-(5-fluoropentyl)-1H-indol-3-yl]tricyclo[3.3.1.1 ^{3,7}]dec-1-yl-methanone; 5F-AB-001; AM2201 adamantyl analog	C24H30FNO	367.51	232	144	233	15.37	show	VBLMLFXPGHZFFP-UHFFFAQYSA-N
Cannabinoids	LTI-701		1-(5-fluoropentyl)-N-phenyl-1H-indole-3-carboxamide		C20H21FN2O	324.40	232	144	233	14.05	show	FZYJALPVILNXEIP-UHFFFAQYSA-N
Cannabinoids	LTI-701		1-(5-fluoropentyl)-N-phenyl-1H-indole-3-carboxamide		C20H21FN2O	324.40	232	144	233	14.01	show	FZYJALPVILNXEIP-UHFFFAQYSA-N

Remark: If the molecular weight and/or empirical formula are known the number of possible candidates can be reduced. For molecular weight search we recommend to use the whole number (neglect decimal digits).

Example 3: You have formal name of a compound as for example:

1-[(tetrahydro-2H-pyran-4-yl)methyl]-N-tricyclo[3.3.1.1^{3,7}]dec-1-yl-1H-indazole-3-carboxamide

Does the compound exist in the RESPONSE database?

First step: Apply the “search field” and search for: *1-[(tetrahydro-2H-pyran-4-yl)methyl]-N-tricyclo[3.3.1.1^{3,7}]dec-1-yl-1H-indazole-3-carboxamide*

There is no match in the database. Is this correct? Go to the next step and try another way around.

Second step: Go to the internet, apply OPSIN application (<http://opsin.ch.cam.ac.uk/>) and generate the StdInChI Key from the IUPAC name.


The screenshot shows the OPSIN web application interface. At the top, there are logos for the University of Cambridge and the Centre for Molecular Informatics. The main heading is "OPSIN: Open Parser for Systematic IUPAC nomenclature". Below this, there is a navigation breadcrumb: "University of Cambridge > Department of Chemistry > Centre for Molecular Informatics". A search input field contains the IUPAC name: "1-[(tetrahydro-2H-pyran-4-yl)methyl]-N-tricyclo[3.3.1.1^{3,7}]dec-1-yl-1H-indazole-3-carboxamide". A "Submit" button is to the right of the input field. Below the input field, there is a message: "Updated 16/10/16: OPSIN 2.2.0 has been released! This version is available from BitBucket and Maven central. If you have found OPSIN useful in your work citing our paper would be very much appreciated. Depiction courtesy of the Indigo Toolkit". A chemical structure is displayed, showing a tetrahydro-2H-pyran ring connected to an indazole ring, which is further connected to a tricyclo[3.3.1.1^{3,7}]decane ring. Below the structure, the "StdInChI:" label is followed by the InChI string: "InChI=1S/C24H31N3O2/c28-23(25-24-12-17-9-18(13-24))11-19(10-17)14-24)22-20-3-1-2-4-21(20)27(26-22)15-16-5-7-29-8-6-16/h1-4,16-19H,5-15H2,(H,25,28)". The "StdInChIKey:" label is followed by the key: "WFDRVDUEMGJILX-UHFFFAOYSA-N". A link "(Click to search the internet for this structure)" is provided. Below the key, the "SMILES:" label is followed by the SMILES string: "O1CCC(CC1)CN1N=C(C2=CC=CC=C2)C(=O)NC12CC3CC(CC(C1)C3)C2".

Figure 7: StdInChIKey from formal name

Third step: Copy the generated code (*WFDRVDUEMGJILX-UHFFFAOYSA-N*) into the RESPONSE database “Search field”.

Result: ADAMANTYL-THPINACA (see Figure 8).

Show entries

Substance class	substance (NPS) common name	structure (created by OPSIN free tool)	NPS1 systematic name	other names	Formula per base form	Mw (g/mol) per base form NPS1	MS (BP1)	MS (BP2)	MS (BP3)	GC-MS-RT NFL/min	MS spectrum (picture)	StdIn ChIKey
Cannabinoids	ADAMANTYL-THPINACA		N-(1-adamantyl)-1-(tetrahydropyran-4-ylmethyl)indazole-3-carboxamide		C24H31N3O2	393.52	135	243	145	17.52	show	VFDRVDUEMGJLLX-UHFFFAOYSA-N

Showing 1 to 1 of 1 entries (filtered from 422 total entries)
 Table updated 28.10.2016
 19748 hits

Figure 8: Search along StdInChIKey returns compound ADAMANTYL THPINACA.

DRAFT

4. Substance descriptors: structure, common name, systematic name, standard InChI key and consistency of systematic chemical identifiers within and between databases

Each distinct compound has a unique molecular structure which can be designated by a structural formula. *Therefore, the expectation is that each structure will be “accompanied” by characteristic and unique name and/or other unique descriptor recognizable by humans and by computerized systems and different types of electronic databases, either local or public opened. The unique descriptor – substance identifier enables consistency of substance coding within and between databases.*

The biggest challenge was to find the most robust “unique” structure descriptor. Common names and synonyms are not unique and mostly not related to the compound structure. Therefore, they are “poor descriptors” of the structure but widely used in conversations. It is known from literature⁵ that some authority-assigned chemical identifiers, which are produced by assignment made by database keepers like Chemical Abstracts Service (CAS) Registry numbers by the American Chemical Society⁶, EC numbers from the European Community Inventory⁷, CID and SID numbers assigned by PubChem⁸, ChemSpider⁹, etc are not the best solution. Namely, different databases use different codes for the same structure and for many novel NPS such codes do not exist at all. Well, some converters¹⁰ can be helpful for finding links of between databases encoding systems.

The alternative to authority-assigned codes and common names are structure-based chemical identifiers. These are derived from structural formula. Anyhow, systematic names (IUPAC based) generated for the same structure often differ. SMILES (simplified molecular-input line-entry system)¹¹ generated for the same structure by different encoding tools shows great diversity. A review of these and other structure notations is given in work of Warr¹². It seems that currently the best way to annotate the structure uniquely is it's encoding by the **International Chemical Identifier** code, i. e. **InChI** string or by its hashed version **InChI Key**. In-depth discussion, some advantages and/or disadvantages one can find in reference⁵ and references listed there.

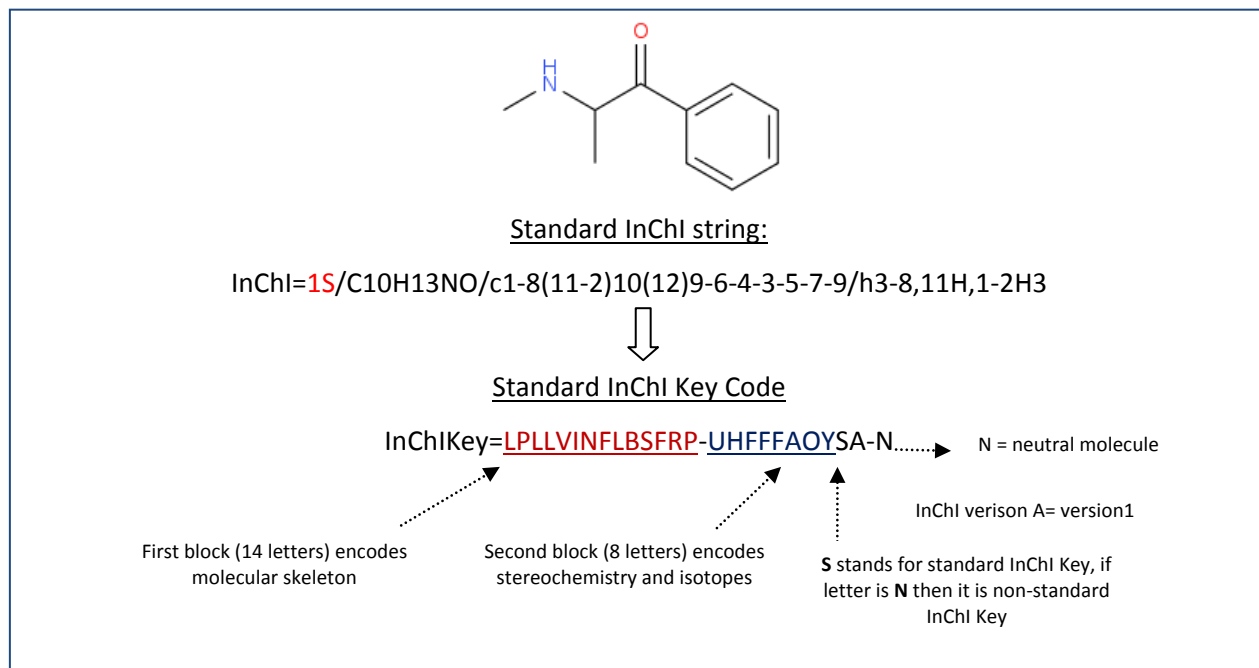


Figure 9: Standard InChI string and corresponding standard InChI Key – example for methcathinone (2-(methylamino)-1-phenyl-propan-1-one).

4.1 RESPONSE project conversions of structure to systematic name and opposite, standard InChI Key and cross validations strategy

During the RESPONSE project database development we explored several types of structure descriptors and finally implemented the **standard InChI Key**¹ as the “unique” one. The main advantage of InChI Key over the InChI string is its uniform length (27 characters – see Figure 9) for any structure, while the string code is longer, of flexible length and contains some symbols not suited the best for search engines and web browsers.

In the RESPONSE project database **standard InChI Key** is always given:

- for the **base form** of compound
- for un-charged state of compound
- for test purchased, collected and seized samples the stereochemistry is not taken into account (i.e. it is undefined or unknown), while for reference materials data from certificates are followed.

One should be aware that codes for base and salt forms, stereoisomers, tautomers and charged species differ!

In the RESPONSE project we explored several free tools for the conversion of structure (in the case of seized, collected and test purchased samples) into IUPAC based name or opposite (when reference materials were in question) and for generation of standard InChIKey code and for cross-validation validation of these data as well. The process is schematically depicted in figure below.

¹ Note: Be aware that nonstandard encoding of structure is also an option. Information on the type of encoding is given as a part of the string/Key- see comments in Figure 9.

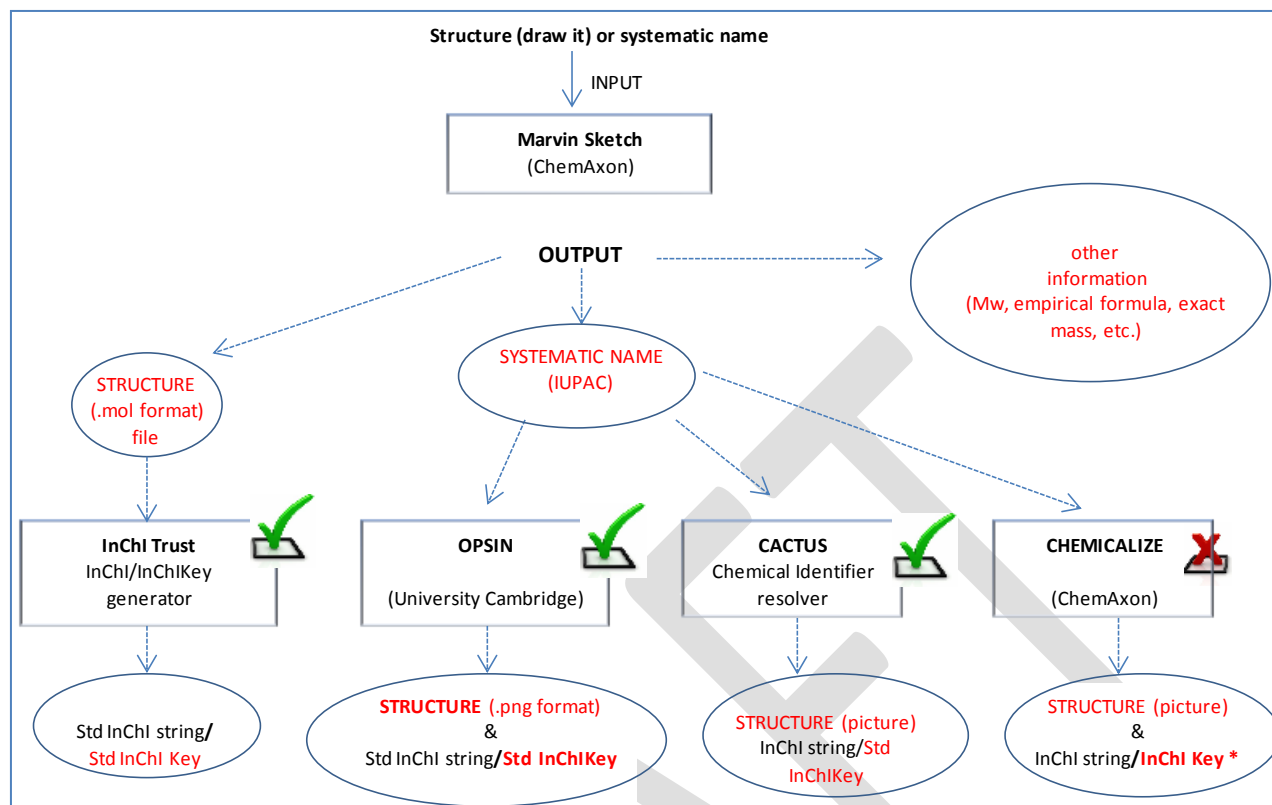


Figure 10: Schematic presentation of RESPONSE project cross-validation tools explored. (CHEMICALIZE * was omitted from the finally implemented system)

*Marvin Sketch*¹³ (ChemAxon) was our basic tool for structural formula generation followed by saving the structure as a electronic "molfile" (readable by other applications), generation of systematic IUPAC and/or traditional name, Mw, exact mass and some other calculations. In addition for cross validations we explored and combined the following free tools:

*INCHI-TRUST*¹⁴ standard and non-standard InChI/InChIKey generator (free software) which is available here: <http://www.inchi-trust.org/downloads/>. This software generates codes from **structural formula** (input is .mol file), which, in our opinion, is the correct way of code generation. So far, we applied the tool on more than 400 entries for cross validation of standard InChIKey obtained from the OPSIN application. We did not find any inconsistencies. Anyhow, the advantage of INCHI-TRUST software is that user can define InChI processing options¹⁵.

*OPSIN*¹⁶ (Open Praser for Systematic IUPAC Nomenclature, Cambridge University) was applied to generate a structure and a standard InChI/InChIKey from systematic name. Structure obtained from

OPSIN was compared with the one from Marvin Sketch. Picture of structure in ".png" format and generated StdInChIKey were implemented into RESPONSE database fields.

CACTUS (NCI/CADD Group)¹⁷ Chemical identifier resolver, which allows the conversion of a given structure identifier into another representation or structure identifier is available here:

<https://cactus.nci.nih.gov/chemical/structure>. This application was applied for cross validation of Standard InChI Key generated by OPSIN.

CHEMICALIZE¹⁸ (ChemAxon) is a powerful online platform for chemical calculations, search and text processing. In the "Calculate" module it generates structure from name and/or name from structure, InChI/InChIKey and performs many other calculations: <https://chemicalize.com/#/calculation>. We applied this tool for different purposes. However, with regard to InChI key we observed a serious "drawback". Namely, in some occasions software generates standard InChI string and/or Key and sometimes nonstandard InChI Key. Along the best of our knowledge user has no influence on this. From the encoding result it can be deduced which type of identifier was actually generated (see Figure 11). Anyhow, for consistency within and between databases such "random" code generation does not fit. This was the main reason why we ruled out this tool (or MarvinSketch directly) for InChI Key generation.

Names and identifiers		A
Common names	methcathinone	
IUPAC name	2-(methylamino)-1-phenylpropan-1-one	
SMILES	<chem>CNC(C)C(=O)C1=CC=CC=C1</chem>	
InChI	InChI=1/C10H13NO/c1-8(11-2)10(12)9-6-4-3-5-7-9/h3-8,11H,1-2H3	
InChIKey	InChIKey=LPLLVINFLBSFRP-UHFFFAOYNA-N	
CAS	28521-94-0,5650-44-2	

Names and identifiers		B
Common names	N/A	
IUPAC name	1-pentyl-N-phenyl-1H-indole-3-carboxamide	
SMILES	<chem>CCCCCN1C=C(C(=O)NC2=CC=CC=C2)C2=CC=CC=C12</chem>	
InChI	InChI=1S/C20H22N2O/c1-2-3-9-14-22-15-18(17-12-7-8-13-19(17)22)20(23)21-16-10-5-4-6-11-16/h4-8,10-13,15H,2-3,9,14H2,1H3,(H,21,23)	
InChIKey	InChIKey=SPUQBSMPSCYQLS-UHFFFAOYSA-N	
CAS	N/A	

Figure 11: Chemicalize outputs for two compounds: A) for methcathinone (2-(methylamino)-1-phenyl-propan-1-one) **non standard** InChI string and Key were generated; B9) for SDB-006-N-phenyl-analog (1-pentyl-N-phenyl-1H-indole-3-carboxamide) **standard** forms of string and Key were the output.

5. References

- ¹ <http://www.policija.si/eng/index.php/generalpolicedirectorate/1669-nfl-page-response>
- ² http://www.policija.si/apps/nfl_response_web/seznam.php
- ³ <http://www.emcdda.europa.eu/>
- ⁴ https://en.wikipedia.org/wiki/International_Chemical_Identifier
- ⁵ Stephen R Heller, Alan McNaught, Igor Pletnev, Stephen Stein and Dmitrii Tchekhovskoi, InChI, the IUPAC International Chemical Identifier, Journal of Cheminformatics 2015, 7:23, DOI: 10.1186/s13321-015-0068-4; <https://jcheminf.springeropen.com/articles/10.1186/s13321-015-0068-4>
- ⁶ CAS REGISTRY and CAS RN FAQs. <http://www.cas.org/content/chemical-substances/faqs>
- ⁷ EC Inventory - ECHA. <http://echa.europa.eu/in/information-on-chemicals/ec-inventory>
- ⁸ <https://pubchem.ncbi.nlm.nih.gov/help.html>
- ⁹ ChemSpider | Search and share chemistry. <http://www.chemspider.com/>
- ¹⁰ <http://cts.fiehnlab.ucdavis.edu/>
- ¹¹ https://en.wikipedia.org/wiki/Simplified_molecular-input_line-entry_system
- ¹² Wendy Warr, Representation of chemical structures. Wiley Interdiscip Rev Comput Mol Sci. 2011;1:557–79, DOI: 10.1002/wcms.36
- ¹³ https://www.chemaxon.com/products/marvin/marvinsketch/?gclid=CjwKEAiAgavBBRCA7ZbggrLSkUcSJACW DexAN5B9vnh01oHXQKgvUvcMGZY065FgDNVurKrfI9VWxoC-sjw_wcB
- ¹⁴ <http://www.inchi-trust.org/>

¹⁵ IUPAC International Chemical Identifier (InChI) Programs InChI version 1, software version 1.04 (September 2011) User's Guide, download able here: <http://www.inchi-trust.org/downloads/> (*INCHI-1-DOC.zip*)

¹⁶ <http://opsin.ch.cam.ac.uk/>

¹⁷ <https://cactus.nci.nih.gov/>

¹⁸ <https://chemicalize.com/welcome>

DRAFT